

技術資料 Report

統計翻訳における人手で作成された大規模フレーズテーブルの効果
SMT with handmade Phrase Table

村上仁一 ^{*1} Jin'ichi Murakami 鏡味良太 ^{*1} Ryouuta Kagami 徳久雅人 ^{*1} Masato
Tokuhisa 池原悟 ^{*1} Satoru Ikehara

^{*1} 〒 680-8552 鳥取市湖山町 4-101 , 鳥取大学工学部知能情報工学科

^{*1}Faculty of Engineering, Tottori University , Minami 4-101,Koyama, Tottori,
680-8552, Japan

キーワード

統計翻訳 , Phrase Table, 手作業

Keyword

SMT, Phrase Table, Handmade

Summary

Recently, the statistical machine translation (SMT) method is very popular for machine translation. This SMT method uses an automatically calculated translation model and language model for large translation pair sentences. The translation model provides the probability that the foreign string is the translation of the native string and is normally controlled using a phrase table.

However, the phrase table is automatically made; it has high coverage but low reliability. On the other side, there are many translation word pairs made by hand, especially in Japanese English translation. These translation word pairs have low coverage but high reliability. Therefore, we added these handmade translation word pairs into the automatically made phrase table.

In this paper, we used 130,000 translation word pairs and the phrase table with added word pairs. As a result of the experiments, we obtained a BLUE score of 13.4% for simple sentences and 8.5% for complex sentences. On the other side, with the base line system, the score was 12.5% for simple sentences and 7.7% for complex sentences. We also studied an ABX test. In simple sentences, 5 sentences were good using the base line, and 23 sentences were good using the proposed method. In complex sentences, 15 sentences were good using the base line, and 35 sentences were good using the proposed method.

As a result of these experiments, the effectiveness of the proposed method was shown.

概要

現在，機械翻訳システムの分野において，対訳データから自動的に翻訳モデルと言語モデルを獲得し，翻訳を行う統計翻訳が注目されている．翻訳モデルでは，原言語の単語列から目的言語の単語列への翻訳を，フレーズテーブルで管理する．しかしフレーズテーブルはプログラムで自動作成するため，カバー率は高いが信頼性は低いと考えられる．一方，手作業で作成した翻訳対は，信頼性は高いがカバー率は低いと考えられる．そこで，それぞれの長所を生かすために，プログラムで自動作成したフレーズ対に手作業で作成した翻訳対を追加することで翻訳精度が向上すると考えた．

実験では，手作業で作成された約 13 万の翻訳対に翻訳確率を与え，プログラムで自動作成したフレーズテーブルに追加した．翻訳実験の結果，BLEU スコアが，日英翻訳の単文では 0.9%，重複文では 0.8% 向上した．また人間による対比較実験を行ったところ，有効性が確認された．

以上の結果から，統計翻訳において手作業で作成した翻訳対を追加する提案手法は有効であることが示された．

1 はじめに

現在，機械翻訳システムの分野において，対訳データから自動的に翻訳モデルと言語モデルを獲得し統計的に翻訳を行う，統計翻訳が注目されている．翻訳モデルは，原言語の単語列から目的言語の単語列への翻訳を確率的に表現するモデルである．言語モデルは，目的言語の単語列に対して，それらが起こる確率を与えるモデルである．翻訳モデルには，大きくわけて語に基づく翻訳モデルと句に基づく翻訳モデルがある．初期の統計翻訳は，語に基づく翻訳モデルであった．語に基づく翻訳モデルでは，原言語の単語から目的言語の単語の対応表を作成する．対応する単語が無い場合は NULL MODEL に対応させる [1]．しかし，翻訳文を生成する時，NULL MODEL に対して，全ての単語の出現を仮定する必要がある．これが翻訳精度が低下する原因の一つになっていた．そのため現在では句に基づく翻訳モデルが主流になっている [2]．

句に基づく翻訳モデルは，原言語の単語列から目的言語の単語列の翻訳に対して確率を付与する．また，NULL MODEL は使用しない．そして，原言語の単語列から目的言語の単語列への翻訳を，フレーズテーブルで管理する．しかし，フレーズテーブルのフレーズ対はヒューリスティックを用いて自動作成されるため，一般にカバー率は高いが信頼性は低いと考えられる．また，フレーズテーブルのフレーズ対は，確率値の信頼性を高めるため，短いフレーズ対に分割される．そのため，長いフレーズ対は少ない．

ところで，日英翻訳では，過去に手作業で作成した日本語の単語列から英語の単語列への翻訳対が大量に作成されている．この翻訳対の信頼性は高いと考えられる．しかし自動作成されたフレーズ対と比較すると，カバー率は低い．そこで，本研究では，それぞれの長所を生かすために，プログラムで自動作成したフレーズ対に手作業で作成された翻訳対を追加することで翻訳精度の向上を目指した．

本研究では，手作業で作成した原言語の単語列から目的言語の単語列への翻訳対を，自動的に作成したフレーズテーブルに追加する．この追加されたフレーズテーブルを利用して日英翻訳の精度向上を試みる．実験では，日英重複文文型パターン辞書 [3] の対訳文対から得られた翻訳対を利用する．手作業で作成された約 13 万の翻訳対に翻訳確率を与え，プログラムで自動作成したフレーズテーブルに追加する．この結果，BLEU スコアが，単文では 12.5% から 13.4% に 0.9% 向上した．また重複文では 7.7% から 8.5% に 0.8% 向上した．また得られた英文 100 文に対し，人間による対比較実験を行ったところ，単文では，従来法が 5 文であるのに対し提案法では 23 文，また重複文では，従来法が 15 文であるのに対し提案法では 35 文，翻訳精度が良いと判断された．

これらの結果から，自動作成されたフレーズテーブルに手作業で作成された翻訳対を追加する，提案手法の有効性が示された．

2 統計翻訳システム

2.1 基本概念

日英の統計翻訳は，日本語文 j が与えられたとき，全ての組合せの中から確率が最大になる英語文 \hat{e} を探索することで翻訳を行う [1]．以下にその基本式を示す．

$$\hat{e} = \operatorname{argmax}_e P(j|e)P(e)$$

$P(e|j)$ は翻訳モデル， $P(e)$ は言語モデルと呼ぶ．翻訳モデルは日本語と英語が対になった対訳コーパスから学習して作成する．また，言語モデルは，出力文側の言語である英語コーパスから学習して作成する．デコーダは言語モデルと翻訳モデルを用いて，尤度の最も高い英文を生成する．

2.2 翻訳モデル

翻訳モデルは，日本語の単語列から英語の単語列または英語の単語列から日本語の単語列へ，確率的に翻訳を行うモデルである．翻訳モデルには，大きくわけて語に基づく翻訳モデルと句に基づく翻訳モデルがある．初期の統計翻訳では，語に基づく翻訳モデルを用いていたが，現在は句に基づく翻訳モデルが翻訳精度が高いため主流になっている．句に基づく翻訳モデルでは，日本語や英語の単語列と確率は，フレーズテーブルで管理される [4]．表 1 にフレーズテーブルの例を示す．

表 1: フレーズテーブルの例

このあたり		this neighborhood		0.027027	0.011973	0.05	0.0087394
最終試験		the final examination		0.125	0.209452	0.0833333	0.0112749
このトーナメント		this tournament		0.0625	0.1842	0.2	0.436971
暖かい冬		a mild winter		0.0714286	0.116379	0.125	0.00362893
このテーブルの位置		the position of this table		0.5	0.00473631	0.5	0.00114018

このテーブルは，左から，“日本語フレーズ”，“英語フレーズ”，“フレーズの英日翻訳確率 $P(j|e)$ ”，“英日方向の単語の翻訳確率の積”，“フレーズの日英翻訳確率 $P(e|j)$ ”，“日英方向の単語の翻訳確率の積”である．

2.3 フレーズテーブルの作成法

句に基づく翻訳モデルは，原言語の単語列から目的言語の単語列の翻訳に対して確率を付与する．これをフレーズテーブルで管理する．以下に作成手順について説明する．

手順 1 単語 alignment の計算 (日英, 英日)

まず, IBM モデル [1] を利用することで, 単語 alignment を得る. これを英日, 日英の両方向に対して行う. つまり, 学習データに対して, 英日方向の単語 alignment と日英方向の単語 alignment を計算する. この tool として GIZA++ [5] が用いられる.

手順 2 単語列 alignment の計算 (union と intersection)

次に, 英日・日英両方向の単語 alignment から, 英日・日英両方向に 1 対多の対応を認めた単語列 alignment を求める. この単語列 alignment は英日・日英両方向の単語対応の積集合 (intersection) と和集合 (union) を利用してヒューリスティックスで求める [6]. 尚, 積集合 (intersection) は, 両方向ともに単語 alignment が存在する場合のみ単語列 alignment を残し, 和集合 (union) は, 少なくとも片方向に単語 alignment が存在する場合に単語列 alignment を残す. 対称な単語列対応を求めるヒューリスティックス (grow-diag-final) は, まず積集合から始まり, 和集合にしかない単語対応が妥当であるかを判断しながら, 単語対応を徐々に加える [7]. なお通常の統計翻訳では, grow-diag-final が利用されている.

手順 3 フレーズテーブルの抽出

単語列 alignment から, ヒューリスティックを用いて日本語単語列と英語単語列のフレーズ対を得る. そのフレーズ対に対して翻訳確率を計算してフレーズテーブルを作成する. 表 2 を学習データとしたとき, grow-diag-final で作成されたフレーズテーブルを表 3 に示す. また, intersection で作成されたフレーズテーブルを表 4 に示す.

表 2: 対訳文の例

日本語	“年 が 明 け た ら 成 人 に な る 。”
英語	“i will come of age after the new year begins .”

表 3: grow-diag-final で作成されたフレーズテーブル (全 12 フレーズ)

が of 1 1 1 1
が 明け of age 1 1 1 1
が 明け たら of age after 1 1 1 1
が 明け たら 成人 come of age after 1 1 1 1
が 明け たら 成人 に will come of age after 1 1 1 1
たら after 1 1 1 1
に will 1 1 1 1
年 が 明け たら 成人 になる 。 i will come of age after the new year begins . 1 0.0987654 1 9.28785e-05
成人 come 1 1 1 1
成人 に will come 1 1 1 1
明け age 1 1 1 1
明け たら age after 1 1 1 1

表 4: intersection で作成したフレーズテーブルの例 (全 185 フレーズから一部抜粋)

年 が 明け たら 成人 に i will come of age after the 0.333333 0.000771606 0.1 1.69351e-05
年 が 明け たら 成人 に i will come of age after 0.333333 0.000771606 0.1 0.000152416
年 が 明け たら 成人 に i will come of age 0.333333 0.000771606 0.1 0.00137174
年 が 明け たら 成人 に i will come of 0.333333 0.000771606 0.1 0.0123457
成人 になる will come of age after 0.0666667 0.0277778 0.1 0.000152416
成人 になる will come of age 0.0666667 0.0277778 0.1 0.00137174
成人 になる will come of 0.0666667 0.0277778 0.1 0.0123457
成人 になる will come 0.0666667 0.0277778 0.1 0.111111

パラメータ intersection で作成したフレーズテーブルは，多くのフレーズ対を持ち，かつ長いフレーズ対を含むことが分かる．

2.4 言語モデル

言語モデルは，目的言語の単語列に対して，それらが起こる確率を与えるモデルである．日英翻訳では，より英語らしい文に対して高い確率を与えることで，翻

訳モデルで翻訳された訳文候補の中から英語として自然な文を選出する。言語モデルとしては N -gram モデルが代表的である。

尚、学習データに表れない単語連鎖確率値を 0.0 とすると、テストデータにおいて、目的言語の全ての単語列の確率が 0.0 になって、単語列が出力されないことがある。そのため、学習データに存在しない単語連鎖確率は、スムージングによって 0.0 以外の確率を割り当てる。代表的なスムージング法として、Backoff や Kneser-Ney がある。これらは高次の N -gram に、低次の N -gram と閾値を掛けて利用する。

2.5 デコーダ

デコーダは翻訳モデルと言語モデルの確率が最大となる文を探索し、出力する。デコーダとして `moses`[4] が代表的である。`moses` はいくつかのパラメータを設定することが出来る。`moses` で設定できるパラメータの例を以下に示す。

- `weight-l` … 言語モデルの重み
- `weight-t` … 翻訳モデルの重み
- `weight-d` … 単語の移動の距離の重み
- `weight-w` … 目的言語の長さに関するペナルティ
- `distortion-limit` … フレーズの並び変えの範囲の制限値

これらのパラメータは、パラメータチューニング (2.6 節) を行うことで最適値を求めることが出来る。

2.6 パラメータチューニング

正解がある `development` データに対して評価値を最大にするように、デコーダのパラメータを最適化することができる。これをパラメータチューニングと呼ぶ。この方法として、Minimum Error Rate Training(MERT)[8] が一般的によく利用される。MERT は、`development` データの、各文について上位 N 個 (通常 100 個) の翻訳候補を出力し、目的の評価値 (通常 BLEU) を最大にするようにデコーダのパラメータの値を調節する。

通常、パラメータチューニングをおこなうと、テストデータの BLEU スコアは上昇する。しかし、実験条件を変更するたびに、パラメータチューニングをおこなうと、多くの時間がかかる。また、本研究では、全ての実験において、実験条件を同一にする必要がある。そのため、パラメータの最適化は行わない。

3 自動的に作成したフレーズテーブルへの翻訳対の追加（提案方法）

3.1 翻訳対への翻訳確率の付与

手作業で作成された翻訳対を，自動的に作成したフレーズテーブルに追加するために，翻訳対に翻訳確率を付与する必要がある．この翻訳確率として，自動作成したフレーズテーブルの翻訳確率を利用する．ただし，フレーズテーブルを作成するときにパラメータ `grow-diag-final` を用いると，確率が付与される翻訳対は少ない．そこで，翻訳確率を与えるためのフレーズテーブルには，多くのフレーズ対を作成するパラメータ `intersection` を用いて作成する．

3.2 翻訳対の追加手順

手作業で作成した翻訳対をフレーズテーブルに追加する手順を図 1 に示す．

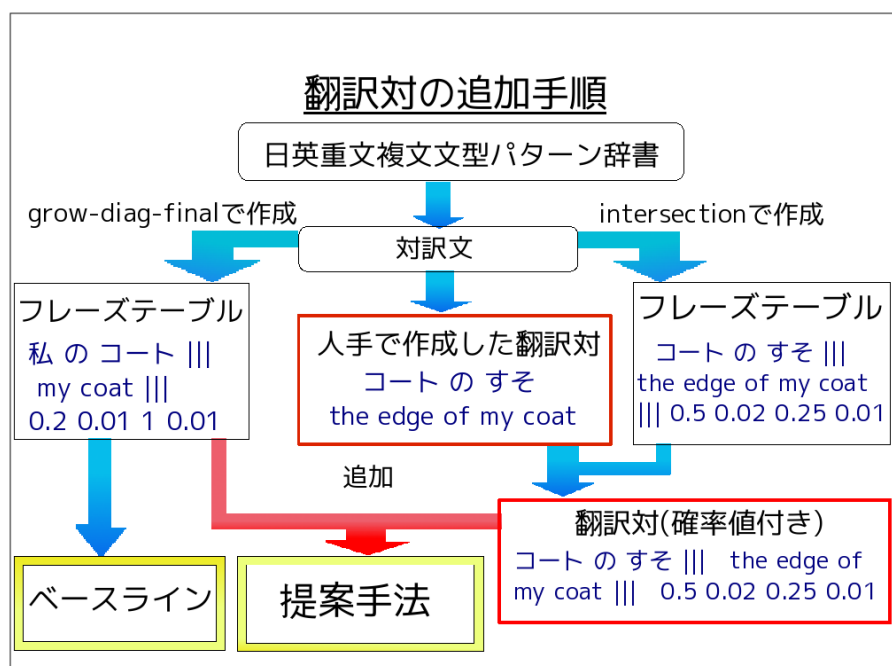


図 1: 手作業で作成したフレーズ対への確率値の付与方法

手作業で作成した翻訳対をフレーズテーブルに追加する手順を以下に示す．

手順 1 前処理

日英重複文型パターン辞書 [3] から対訳文を抽出し “chasen[9]” で形態素解析を行う．英語文に対しては大文字の小文字化を行う．また，句読点の前にスペースを入れる．前処理を行った後の対訳文の具体例を表 8 に示す．

手順 2 intersection を用いたフレーズテーブルの作成

手順 1 で抽出した対訳文を用いてフレーズテーブルを作成する．作成時のパラメータには intersection を用いる．作成したフレーズテーブルの例を表 5 に示す．

表 5: パラメータ intersection で作成したフレーズテーブルの例

オーバー コートを脱ぎ捨て	flung his coat off	0.5	3.88199e-08	0.5	6.46865e-06
コートのすそ	the edge of my coat	0.5	0.000296246	0.166667	5.47368e-13
帽子とコートをかけ	hung up his hat and coat	0.0416667	6.7781e-07	0.0833333	4.50333e-11

手順 3 手作業で作成した翻訳対への翻訳確率値を付与

手順 2 で作成したフレーズテーブルを参照して，手作業で作成した翻訳対に翻訳確率値を付与する．翻訳対が “オーバー コートを脱ぎ捨て flung his coat off” の場合は 1 行目のフレーズ対の翻訳確率値 “0.5 3.88199e-08 0.5 6.46865e-06 ” を付与する．

手順 4 grow-diag-final をもちいたフレーズテーブルの作成

手順 1 で抽出した対訳文を用いてフレーズテーブルを作成する．作成時のパラメータには grow-diag-final を用いる．作成したフレーズテーブルの例を表 6 に示す．

表 6: パラメータ grow-diag-final で作成したフレーズテーブルの例

にオーバー コートを	coat off in	1	0.000479567	1	0.000158639
オーバー コートを	coat off	0.5	0.00153186	1	0.00122977
コート	a coat	0.25	0.322034	0.0333333	0.197333

手順 5 フレーズテーブルの追加

手順 4 で作成したフレーズテーブルに，手順 3 で作成した翻訳確率を付与した翻訳対を追加する．

表 7: 確率値を付与した翻訳対を追加したフレーズテーブルの例

に オーバー コートを		coat off in		1	0.000479567	1			
オーバー コートを		coat off		0.5	0.00153186	1	0.00122977		
オーバー コートを脱ぎ捨て		flung his coat off		0.5	3.88199e-08	0.5	6.46865e-06		
コート		a coat		0.25	0.322034	0.0333333	0.197333		
コートのすそ		the edge of my coat		0.5	0.000296246	0.166667	5.47368e-13		
帽子とコートをかけ		hung up his hat and coat		0.0416667	6.7781e-07	0.0833333	4.50333e-11		

尚，本稿では，手順 4 で作成したフレーズテーブルを用いた翻訳をベースラインと呼び，手順 5 のフレーズテーブルを用いた翻訳を提案手法と呼ぶ．

4 翻訳実験

翻訳実験は，単文と重複文の 2 種類で行う．

4.1 学習データ

単文の翻訳実験には，電子辞書などから抽出した単文 10 万文対 [11] を学習データとして用いる．重複文の翻訳実験には日英重複文文型パターン辞書 [3] から抽出した対訳文対，121,913 文対を用いる．尚，単文 10 万文は，日本語が単文であるが，対訳英文は単文とは限らず複文の場合もある．重複文 121,913 文は，日本文が重文もしくは複文であるが，英文は複文とは限らず単文である場合もある．

前処理 [手順 1] を行った対訳文の例を表 8 に示す．

表 8: 対訳文の例

単文の例
教習所は自動車の運転法を教える。 driving schools teach people how to drive a car .
彼は彼女の甘い言葉に乗った。 he was fooled completely by her sweet words .
若い兵士は彼女の写真を胸に、戦地に向かった。 the young soldier headed for the front with her photograph at his chest .
重複文の例
愛と幸せに満ちた家庭にしたいと思います。 we hope we will make our home full of love and happiness .
彼はときどきうなずきながらわたしの話を聞いた。 he listened to my story nodding his head occasionally .
きみのあいまいな説明ではあの娘がだれだかわからなかった。 i could not recognize the girl from the vague description of her you gave me .

4.2 手作業で作成された翻訳対

手作業で作成された翻訳対は、日英重複文文型パターン辞書 [3] から抽出した対訳コーパスから作成された翻訳対 261,453 個を用いる。この翻訳対は、プロの翻訳者が手動で作成した対訳対で、単語、句、節の単位で対応づけられている。また、この翻訳対は日本語文が重複文で英語が単文もしくは重複文である対訳コーパスから抽出されている。文献 [10] に、この翻訳対の詳しい説明がある。基本的には、日本語文と英語文の対訳文から日本語パターンと英語パターンを作成する。このとき、作成できる日英翻訳対を利用する。翻訳対の抽出において、長さの制限はおこなっていない。また、重複する句は抽出していない。例を表 9 に示す。

表 9: 翻訳対の作成例

日本語文:	クロッカスは春に咲き、サフランは秋に咲く。
英語文:	Crocus blooms in the spring while saffron blooms in the fall.
日本語パターン:	$N1$ は, $TIME2$ に $V3$ 、 $N4$ は, $TIME5$ に $V6$ 。
英語パターン:	$N1$ in $N2$ while $N4$ $V6$ in $N5$.
日英翻訳対	
$N1$:	クロッカス Crocus blooms
$TIME2$:	春 the spring
$V3$:	咲き blooms
$N4$:	サフラン saffron
$TIME5$:	秋 the fall
$V6$:	咲く blooms

手作業で作成された翻訳対の例を表 10 に示す。翻訳対の分布図を図 2 に示す。この図では、縦軸が全体に占める割合で、横軸が 1 つの翻訳対における単語数である。日本語における単語数を J 、英語における単語数を E で示している。これからわかるように、2 単語のフレーズが最も多く、単語数と、その単語数がしめる割合は、zipf の法則に沿っていることがわかる。なお、本稿では、手作業で作成された単語列の対訳対を翻訳対と呼ぶ。

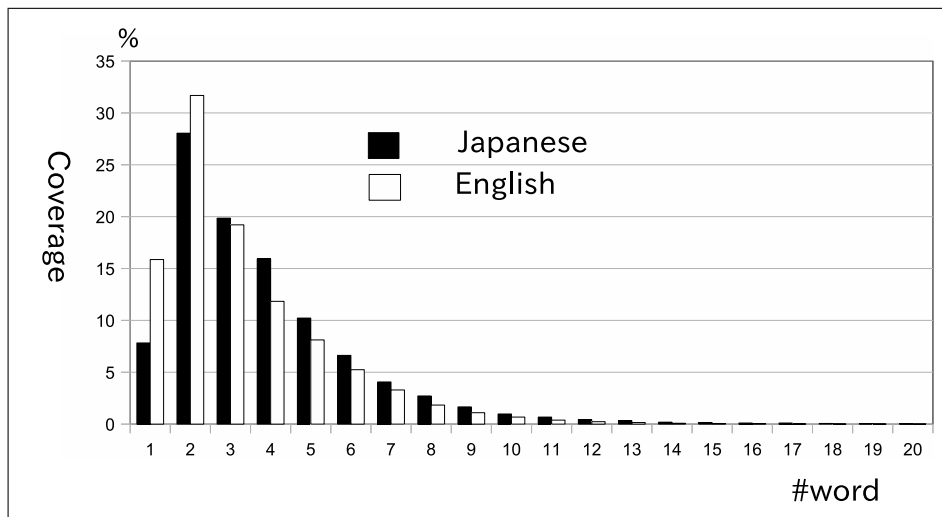


図 2: 手作業で作成した翻訳対の単語数の分布図

表 10: 手作業で作成された翻訳対の例

コート の すそ	the edge of my coat
朝晩 の ラッシュ 時に	during the morning and evening rush hours
国産 の コート は 英国 製 より だいぶ 落ちる	home-made coats are by far inferior to those made in britain

4.3 テストデータ

テストデータには，電子辞書などから抽出した単文 1000 文対 [11] を用いる．重複文の翻訳実験には日英重複文文型パターン辞書 [3] から抽出した対訳文対 1,000 文対を用いる．ただし，テストデータは学習データ（4.1 節）や手作業で作成された翻訳対（4.2 節）と，別の辞書を利用する．従って，テストデータは，学習データや翻訳対に対して open data となる．

4.4 翻訳モデルと言語モデルとデコーダ

1. フレーズテーブルの作成

翻訳モデルはフレーズテーブルで管理される．フレーズテーブルの作成には，train-phrase-model.perl[12] を用いて自動的に作成する．尚，本稿では，プログラムで自動作成した単語列の対訳対をフレーズ対と呼ぶ．また，フレーズ対の最大の単語数を決める max phrase length は 20 とする．

2. N -gram モデルの学習

言語モデルには， N -gram モデルを用いる． N -gram モデルの学習には，“SRILM[13]”を用いる．本研究では 5-gram モデルを用いる．また，スムージングのパラメータには，Kneser-Ney である “-ukndiscount” を用いる．

3. デコーダ

デコーダは “moses[4]” を用いる．また，翻訳モデルには，日英翻訳確率と英日翻訳確率の相互情報を用いる [14]．したがって，翻訳モデルの重み “weight-t” は “0.5 0 0.5 0 0” とする．また，翻訳時にフレーズの位置の変化に柔軟に対応するため，単語の移動重み “weight-d” は 0.2 とする．また単語の移動距離の制限 “distortion-limit” は，-1（無制限を意味）とする．その他は，default 値とする．

4.5 評価方法

評価は、コンピュータによる自動評価と人間による評価の、2種類で行う。

1. 自動評価

機械翻訳システムの翻訳精度を自動評価する手法として、あらかじめ実験者が用意した正解文と、翻訳システムが出力した文とを比較する手法が利用されている。この自動評価法には多くの方法が提案されている。本研究では、 N -gram を用いた BLEU[15] と類似単語辞書を用いた METEOR[16] を用いる。

2. 人間による評価

人間による評価として、対比較実験をおこなう。得られた英文から 100 文をランダムに抽出し、ベースラインの翻訳結果と提案手法の翻訳結果のどちらの翻訳結果が優れているかを人間で判断する。その際、本研究において固有名詞の未知語はローマ字変換して評価し、それ以外の未知語は存在しないとして評価を行う。

4.6 実験結果 フレーズテーブルの増加数

ベースラインのフレーズ数、確率値が付与できた翻訳対の数、最終的に作成されたフレーズ数を表 11 に示す。

表 11: 総フレーズ数

	プログラム作成によるフレーズ数 (ベースライン) (grow-diag-final)	確率のついた翻訳対 (手作業)	追加後のフレーズ数 (提案方法)
単文	572,307	120,792	693,099
重複文	727,848	130,892	858,740

手作業で作成された翻訳対は、261,453 対であった。しかし約半数以上に対して確率値を付与できなくて、削除されていることがわかる。また、提案法におけるフレーズテーブルのフレーズ数は、ベースラインと比較すると約 2 割増加している。確率が付与された翻訳対の例を表 12 に示す。

表 12: 確率値が付与された翻訳対の例

コート の すそ the edge of my coat
0.5 0.00190997 0.25 6.48554e-06
朝晩 の ラッシュ 時 に 電車 に 乗る
take the train during the morning and evening rush hours
0.1 6.68278e-20 0.125 1.83956e-24
国産 の コート は 英国 製 より だいぶ 落ちる
home-made coats are by far inferior to those made in britain
0.05 8.90443e-29 0.0454545 5.07768e-33

4.7 実験結果 翻訳精度の評価

1. 自動評価

日英翻訳のテストデータには，単文 1,000 文と重複文 1,000 文を用いる．ベースラインと提案手法の翻訳精度の自動評価の結果を表 13 に示す．

表 13: 実験結果 (テストデータ 1,000 文)

テスト文	評価法	ベースライン	提案手法
単文	BLEU	0.125	0.130
	METEOR	0.389	0.404
重複文	BLEU	0.077	0.080
	METEOR	0.296	0.303

結果から，単文，重複文のいずれの翻訳においても提案手法の翻訳精度が向上していることが分かる．

2. 人間による対比較実験

表 13 の日英翻訳結果からランダムに抽出した 100 文に対して，人間による対比較実験を行う．対比較実験において，ベースラインの翻訳結果の方が，優れていると評価した文を，“提案手法 ×” とする．提案手法の翻訳結果が，ベースラインの翻訳結果より優れていると評価した文を，“提案手法 ” とする．また，ベースラインと提案手法で翻訳結果が変化しなかった文を“変化無し” とする．対比較実験の結果を表 14 に示す．

表 14: 対比較実験の結果

テストデータ	提案手法 ×	提案手法	変化無し
単文	3	21	76
重複文	5	15	80

結果から，全ての翻訳において，提案手法が優れている割合が高くなっていることが分かる．

5 翻訳対の翻訳確率の重みの最適化

4章の実験では，手作業によって作成した翻訳対に，パラメータ intersection で作成した翻訳確率を付与した．しかし，手作業で作成された翻訳対は信頼性が高いと考えられる．そこで，翻訳対に付与する翻訳確率の重みを大きくした方が翻訳精度が向上すると考えられる．そこで，翻訳対に付与する翻訳確率の重みを大きくした実験を行う．

5.1 翻訳確率の重みを変えた翻訳実験

単文および重複文の翻訳実験において，手作業で作成した翻訳対の翻訳確率の重みを2倍，4倍，8倍に変化させたときの，BLEUスコアとMETEORの変化を調査する．結果を表15に示す．

表 15: 翻訳確率の重みを変えたときの翻訳実験結果

テストデータ	評価法	ベースライン	1倍	2倍	4倍	8倍
単文	BLEU	0.125	0.130	0.134	0.130	0.128
	METEOR	0.389	0.404	0.404	0.401	0.402
重複文	BLEU	0.077	0.080	0.081	0.085	0.085
	METEOR	0.296	0.303	0.303	0.308	0.309

日英の翻訳において，単文の翻訳時には翻訳確率の重みを2倍，重複文の翻訳時には8倍にした時に翻訳精度がもっとも良かった．最適な翻訳確率の重みを用いたときの提案手法の翻訳精度と，ベースラインの翻訳精度の差を比較した場合，BLEUでは，単文で0.9%，重複文で0.8%向上していることがわかる．

5.2 対比較実験

表 15 の翻訳結果 100 文に対して、表 14 と同じ条件で対比較実験を行った結果を表 16 に示す。

表 16: 対比較実験の結果

テストデータ	提案手法 ×	提案手法	変化無し
単文	5	23	72
重複文	15	35	50

表 14 と比較すると、特に重複文において改善が見られる。

5.3 対比較実験の解析

表 16 における対比較実験の例文を以下に示す。

1. 提案手法が優れている例

提案手法が優れていると評価した例を表 17 に示す。

表 17: 提案手法が優れていると評価した例

単文
入力文:彼は 苦しい 立場 に ある 。 正解文:he is in a difficult position . ベースライン:he is in 苦しい . 提案手法:he is in a delicate position .
入力文:彼は その 文章 を 全部 暗記 して しまった 。 正解文:he has memorized the entire sentence by heart . ベースライン:he gave all the sentence has 暗記 . 提案手法:he has a memorize all the sentence .
入力文:私 は 見知らぬ 人 から 声 を かけられた 。 正解文:a stranger called out to me . ベースライン:i was submitted a voice from a stranger . 提案手法:i was called from a stranger .
重複文
入力文:ten divided by 2 makes 5 . 正解文:1 0 を 2 で 割れば 5 になる 。 ベースライン:if you get 5 to 10 broken by two . 提案手法:ten divided by two gives five to you .
入力文:背 が 高くない と ファッションモデル に はなれない 。 正解文:you can 't be a fashion model if you are not tall enough . ベースライン:it is hard to ファッションモデル not tall . 提案手法:it is hard to a model not tall .
入力文:あんな 映画、 見 に 行か なければ よかった 。 正解文:i shouldn 't have gone to see such a movie . ベースライン:i see a movie , i should not have to go . 提案手法:we should not have to go see a movie , like that .

2. 提案手法が劣っていると評価した例

提案手法が劣っていると評価した例を表 18 に示す .

表 18: 提案手法が劣ると評価した例

単文
入力文:彼は重い罰を受けた。 正解文:he received severe punishment . ベースライン:he received a heavy penalty . 提案手法:he was severely punished .
入力文:彼は自分を責めさいなんだ。 正解文:he severely reproached himself . ベースライン:he is a 責めさいなん himself . 提案手法:he is 責めさいなん myself .
入力文:雨が降らないだろう。 正解文:It will probably not rain . ベースライン:it will not rain . 提案手法:it rains will not .
重複文
入力文:欲しいだけお金を持って行ってもよい。 正解文:you may go taking as much money as you want . ベースライン:you may have been much money . 提案手法:i want to take as much money may .
入力文:こんなにたくさんの人に見られているのでは緊張してしまうでしょう。 正解文:she may get nervous with such a large audience watching her . ベースライン:so many people are so nervous i will have to be seen . 提案手法:so many people have been so nervous i will have to look at .
入力文:you can get some allowance if you help me clean the house . 正解文:掃除を手伝ってくればおこづかいをあげる。 ベースライン:we will keep you to help me raise pocket money . 提案手法:we will keep spending money to help us to raise .

5.4 パラメータを最適化した翻訳実験

1. パラメータの最適化

通常，統計翻訳においては，翻訳精度の向上を目的として，パラメータの最適化を行う．この節では，パラメータの最適化をおこなったときの，提案方法の有効性を調査する．パラメータの最適化には，Minimum Error Rate

Training(MERT)[8]を用いる。尚、フレーズテーブルの作成には moses に付属している train-factored-phrase-model.perl を用いる。また、reordering モデルも組み込む。

2. development データ

development データは、単文の実験も重複文の実験も、テストデータ(4.3節)と同一の辞書から抽出したデータを利用する。単文の翻訳実験には、development データに単文 100 文を使用してパラメータの最適化を行う。重複文の翻訳実験には development データに重複文 1,000 文を使用してパラメータの最適化を行う。

3. 翻訳実験の結果

翻訳実験の結果を表 19 に示す。

表 19: パラメータチューニングを行った実験結果

テスト文	評価法	ベースライン	提案手法
単文	BLEU	0.131	0.140
	METEOR	0.383	0.395
重複文	BLEU	0.084	0.086
	METEOR	0.313	0.325

表 19 の結果から、パラメータの最適化を行った翻訳実験においても、BLEU が単文において 0.9%、重複文において 0.2% 上昇し、提案手法の有効性が示された。

6 考察

6.1 提案手法の効果の分析

表 5.2 の対比較実験の結果において、翻訳結果が変化した 78 文中、58 文が提案手法が優れていると評価した。この評価の理由として、妥当な語順による向上と未知語の減少に分けることが出来る。以下にその分析結果を述べる。

1. 妥当な語順による向上

提案手法の翻訳結果がベースラインと比較して、妥当な語順となって文質が向上したと判断した例を表 20 に示す。

表 20: 妥当な語順による向上例

単文
入力文:子供達は窓ガラスを割った。 正解文:The children broke the window panes . ベースライン:the children fell below the windows . 提案手法:the children broke the window .
正解文:Next time I will return 2 years time from now . ベースライン:i now 戻る for two years . 提案手法:i will return to the next two years .
正解文:The train will arrive at the station about 4 o'clock . ベースライン:the train at around the station , will reach . 提案手法:the train arrived at the station at four o'clock around .
重複文
入力文:この学校では、毎年五百名の学生が卒業していく。 正解文:five hundred students graduate from this school every year . ベースライン:this is the name of 500 students graduate school year , and go . 提案手法:this school is 500 students graduate of the year , and go .
入力文:友達のお父さんに、駅まで車で送っていただきました。 正解文:the father of my friend drove me to the station . ベースライン:i would like to the station by a friend of mine , to your father . 提案手法:i would like to the station , so send it to my friend 's father .
入力文:彼女の声は電車の音にかき消された。 正解文:her voice was drowned out by the sound of the train . ベースライン:the noise of the train was lost in her voice . 提案手法:her voice was lost in the sound of the train .

2. 未知語の減少

未知語が減少したことにより翻訳精度が向上した例を表 21 に示す。

表 21: 未知語が減少した例

単文
<p>入力文:日本 と 大陸 は かつて つながって いた。</p> <p>正解文:the continent and the island of japan were linked in the old days .</p> <p>ベースライン:he was once a つながっ japan and the continent .</p> <p>提案手法:japan is linked up to the continent was once .</p>
<p>入力文:地球 は 温暖化 しつつ ある。</p> <p>正解文:global warming is developing .</p> <p>ベースライン:the earth is 温暖 they are .</p> <p>提案手法:they are warming the earth .</p>
<p>入力文:彼 は 苦しい 立場 に ある。</p> <p>正解文:he is in a difficult position .</p> <p>ベースライン:he is in 苦しい .</p> <p>提案手法:he is in a delicate position .</p>
重複文
<p>入力文:食事 を 減らせ ば 誰 でも やせる。</p> <p>正解文:you can lose weight by reducing what you eat .</p> <p>ベースライン:減らせ , and everyone やせる .</p> <p>提案手法:if you eat any 減らせ slimming down .</p>
<p>入力文:背 が 高くない と ファッションモデル に はなれない。</p> <p>正解文:you can 't be a fashion model if you are not tall enough .</p> <p>ベースライン:it is hard to ファッションモデル not tall .</p> <p>提案手法:it is hard to a model not tall .</p>
<p>入力文:手当て が 早ければ、 彼は 助かっていた。</p> <p>正解文:he could have been saved if he had been given medical treatment sooner .</p> <p>ベースライン:the 早けれ if he had a narrow escape . 手当て</p> <p>提案手法:the treatment of the 早けれ if he had .</p>

3. 妥当な語順になった文と未知語が減少した文の比較

提案手法が優れていると評価した 58 文において、未知語が減少した文数と、妥当な語順になった文数を表 22 に示す。

表 22: 妥当な語順になった文数と未知語が減少した文数の比較

テストデータ	未知語が減少	妥当な語順になった文
単文	10	13
重複文	4	31

表 22 から，約 8 割の文が，未知語の減少よりも妥当な語順になって翻訳精度が向上していると判断された．つまり，提案手法の有効性は，主に妥当な語順になった文の増加にあると言える．

6.2 今後の課題

今後の課題として，以下の項目がある．

1. 手作業で作成された翻訳対の翻訳確率の最適化

手作業で作成された翻訳対は信頼性が高いため，翻訳確率値が大きい方が，高い翻訳精度が得られると考え，第 5 章において翻訳対に付与した翻訳確率の重みを変化させて翻訳実験を行った．この結果，翻訳精度が向上した (表 15)．しかし，重みを大きすぎると翻訳精度が低下した．この結果から，重みの最適化が必要であると考えている．そして，この重みの最適化に MERT が使用できると考えている．

2. 翻訳確率値を付与できなかった翻訳対の追加

本研究では約 26 万個の手作業で作成された翻訳対のうち，約 13 万個の翻訳対に翻訳確率値を付与できた．そして，翻訳確率値を付与できなかった翻訳対約 13 万個は，削除した．そこで，翻訳確率値を付与できなかった翻訳対約 13 万個に対して，翻訳確率として閾値を与えて，翻訳実験を行った．しかし，どのような閾値を与えても，BLEU, METEOR とともに低下した．今後，確率を付与できなかった翻訳対の，確率の付け方を考えてみたい．

3. 述語節に関する翻訳対の追加

翻訳において，述語節が正しく翻訳されているか否かは，人間の評価において重要な判断要素となりやすい．つまり，述語節が正しく翻訳されると，文の意味が分かりやすくなり，人間による翻訳精度の評価が向上する，そこで，今後は特に，述語節に関する翻訳対を追加し，翻訳精度の調査を行いたいと考えている．また，英辞郎 [17] には，手作業によって作成された 200 万以上の日英の翻訳対がある．これを利用することでさらに翻訳精度が向上すると考えている．

7 おわりに

本研究では，手作業で作成した信頼性の高い翻訳対を，プログラムで自動作成したフレーズテーブルに追加して，単文と重複文における日英翻訳の精度評価を行った．約 13 万の翻訳対を追加し，追加した翻訳対の翻訳確率の重みを変えた結果，BLEU スコアが日英翻訳において，単文では 12.5%から 13.4%に 0.9%向上した．また重複文では 7.7%から 8.5%に 0.8%向上した．また出力英文 100 文に対し人間による対比較実験を行ったところ，単文では，従来法が良いと判断された文が 5 文であるのに対し，提案法では 23 文，また重複文では，従来法が良いと判断された文が 15 文であるのに対し，提案法では 35 文となった．

以上の結果から，提案手法の有効性が示された．今回の実験では日英重複文文型パターン辞書 [3] の対訳文対から，手作業で作成した翻訳対を追加した．今後は他の辞書の翻訳対も追加して，翻訳精度の調査をすることを考えている．また，追加する翻訳対の翻訳確率値に対する重みの最適化の方法についても考えていく．

謝辞

日英重複文文型パターン辞書の対訳文対や，この対訳対から得られる翻訳対の作成には，多くの方の協力を得ました．基本的には鳥バンクの作成において関連した方々です．特に，以下の人に厚くお礼を申し上げます（順不同）．

白井諭，藤波進，小見佳恵，阿部さつき，木村淳子，竹内奈央，小船園望（以上 NTT-AT），池田尚志（岐阜大学），佐良木昌（長崎純心大），新田義彦（日本大学），柴田勝征（福岡大学），山本理恵（鳥取大学工学部：事務局），大山芳史（NTT-CS 研），衛藤純司（ランゲージウェア）

参考文献

- [1] Brown, Peter F., John Cocke, Stephen Della Pietra, Vincent J. Della Pietra, Frederick Jelinek, John D. Lafferty, Robert L. Mercer, and Paul S. Roossin, “A Statistical Approach to Machine Translation”, Computational Linguistics. Vol. 16. No. 2. pp. 7985, 1990.
- [2] Philipp Koehn, Franz J. Och, and Daniel Marcu, “Statistical phrase-based translation”, HLT-NAACL 2003, pp. 127-133, 2003.
- [3] 鳥バンク, “<http://unicorn.ike.tottori-u.ac.jp/toribank/>”, 2007.
- [4] Philipp Koehn, Marcello Federico, Brooke Cowan, Richard Zens, Chris Dyer, Ondrej Bojar, Alexandra Constantin, Evan Herbst, “Moses: Open Source

- Toolkit for Statistical Machine Translation”, Proceedings of the ACL 2007 Demo and Poster Sessions, pages 177-180, 2007.
- [5] Franz Josef Och, Hermann Ney, “A Systematic Comparison of Various Statistical Alignment Models”, Computational Linguistics, volume 29, number 1, pp. 19-51, 2003.
- [6] Franz Josef Och and Hermann Ney, “A systematic comparison of various statistical alignment models”, Computational Linguistics, 29(1), pp. 19-51. 2003.
- [7] 山本 幹雄, 藤井敦, 内山将夫, 宇津呂武仁, “統計的機械翻訳における特許文翻訳に関する講習会”, pp.11 2007.
- [8] Franz Josef Och, “Minimum Error Rate Training in Statistical Machine Translation”, Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics, July 2003, pp. 160-167.
- [9] 松本 裕治, “形態素解析システム「茶筌」”, 情報処理 41(11), 1208-1214, 2000.
- [10] 池原 悟, “非線形言語モデルによる自然言語処理”, 岩波書店, ISBN978-4-00-005882-7, pp.220-242, 2009.
- [11] 村上 仁一, 池原 悟, 徳久 雅人, “日本語英語の文対応の対訳データベースの作成”, 第7回「言語, 認識, 表現」年次研究会, 2002.
- [12] NAACL 2006 WORKSHOP ON STATISTICAL MACHINE TRANSLATION Shared Task, Exploiting Parallel Texts for Statistical Machine Translation Shared Task Baseline System, training-release-1.3.tgz, “<http://www.statmt.org/wmt06/shared-task/baseline.html>”
- [13] Andreas Stolcke, “SRILM - An Extensible Language Modeling Toolkit”, in Proc. Intl. Conf. Spoken Language Processing, Denver, Colorado, September 2002
- [14] Jin’ichi Murakami, Masato Tokuhisa, Satoru Ikehara, “Statistical Machine Translation using Large J/E Parallel Corpus and Long Phrase Tables”, International Workshop on Spoken Language Translation 2007, pp.151-155, 2007.
- [15] K. Papineni, S. Roukos, T. Ward, W. J. Zhu, “BLEU: a method for automatic evaluation of machine translation”, 40th Annual meeting of the Association for Computational Linguistics pp. 311-318, 2002.

- [16] Banerjee, S. and A. Lavie, “METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments”, Proceedings of Workshop on Intrinsic and Extrinsic Evaluation Measures for MT and/or Summarization at the 43th Annual Meeting of the Association of Computational Linguistics (ACL-2005), June 2005.
- [17] EDP 編集, “英辞郎 第4版”, 株式会社アルク, ISBN 4757414560, 2008.
- [18] 鏡味良太, 村上仁一, 徳久雅人, 池原悟, “統計翻訳における人手で作成された大規模フレーズテーブルの効果”, 言語処理学会第15回年次大会, pp.224-227, 2009

略歴

村上仁一: 1984年 筑波大学第3学群基礎工学類卒業. 1986年 筑波大学大学院修士課程理工学研究科理工学専攻修了. 同年 NTT に入社. NTT 情報通信処理研究所に勤務. 1991年 国際通信基礎研究所 (ATR) 自動翻訳電話研究所に出向. 1995年 NTT 情報通信網研究所に復帰. 1997年 豊橋技術科学大学にて博士(工学). 1998年 鳥取大学工学部知能情報工学科に転職. 現在に至る. 主に音声認識のための言語処理の研究を行う. 最近は統計翻訳の研究に従事. 電子情報通信学会, 日本音響学会, 言語処理学会各会員.

鏡味良太: 2009年 鳥取大学工学部知能情報工学科卒業. 同年 エヌデック株式会社に入社. 現在に至る.

徳久雅人: 1995年 九州工業大学大学院情報工学研究科博士前期課程修了. 同年同大学情報工学部知能情報工学科助手, 2002年 鳥取大学工学部知能情報工学科助手, 現在, 同大学大学院工学研究科情報エレクトロニクス専攻講師. 博士(工学). 情報処理学会, 電子情報通信学会, 人工知能学会, 言語処理学会各会員.

池原悟: 1967年 大阪大学基礎工学部電気工学科卒業. 1969年 同大学院修士課程終了. 同年日本電信電話公社に入社. 1982年 情報処理学会論文賞. 1993年 情報処理学会研究賞. 1995年 日本科学技術情報センター賞. 1995年 人工知能学会論文賞. 1996年 スタンフォード大学客員教授. 1996年 鳥取大学工学部教授. 2002年 電気通信普及財団賞受賞. 2006年 文部科学大臣表彰 科学技術賞. 2009年 12月逝去. 工学博士. 数式処理, トラフィック理論, 自然言語処理の研究に従事.